

Comparison Performance of Lymphocyte Classification for Various Datasets using Deep Learning

Syadia Nabilah Mohd Safuan¹, Mohd Razali Md Tomari², Wan Nurshazwani Wan Zakaria³, Nor Surayahani Suriani⁴
Faculty of Electronic and Electrical Engineering, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia^{1,3,4}
Principle Researcher in Research Centre for Applied Electromagnetics, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia²

Abstract—Analyzing and classifying five types of Lymphocyte White Blood Cell (WBC) is important to monitor the lack or excessive amount of cell in human body. These harmful amount of cell must be detected early for the early treatment can be run to the patient. However, the process may be tedious and time consuming as it is done manually by the experts. Other than that, it may yield inaccurate result as it depends on the pathologist skill and experience. This work presents a way that can be the second opinion to the experts using computer aided system as a solution. Convolutional Neural Network (CNN) is applied to the system to avoid complex structure and to eliminate the features extraction process. Three CNN models of mobilenet, resnet and VGG-16 is experimented on three different datasets which are kaggle, LISC and IDB-2. Kaggle, LISC and IDB-2 dataset consist of 6000, 242 and 260 images respectively. The result is divided into two parts which are dataset and model. As for IDB-2 dataset, the best model is VGG with training and validation accuracy of 0.9721 and 0.7913 respectively. While for kaggle and LISC dataset, the best model is resnet as it achieved training accuracy of 0.9713 and 0.9771 respectively. The highest validation accuracy for kaggle is 0.5955 and 0.5781 for LISC. Lastly, the best database that is most suitable for all model is IDB-2 database. It obtained highest training and validation accuracy for all model of mobilenet, resnet and VGG-16.

Keywords—Convolutional neural network; Google colab; training accuracy; validation accuracy; white blood cell

I. INTRODUCTION

White Blood Cell (WBC) is one of other particles in human blood. This cells help to fight disease and virus as its presence help to boost the immune system [1]. However, having extra and unnecessary WBC in the blood could be harmful. This is where the WBC analysis is needed. WBC analysis is undeniably important as it helps to prevent disease's complications and an early prevention can be made [2]. Other than that, it is very helpful for diagnosing the patient's health condition [3].

WBC has five types which are Eosinophil, Basophil, Neutrophil, Lymphocyte and Monocyte [4]. They vary in terms of its shape, number or lobes and sizes of its nucleus and cytoplasm as showed in Fig. 1 [5, 6]. It is also differentiated by the nucleus' stain [7]. The number of each cells must be maintained as excessive number of cell will create problem to the patient's health.

Conventionally, WBC analysis is done manually where the images were placed under the microscope and pathologist analyze it manually [8]. This process takes time and creates confusion as the result is highly dependent on the pathologist's skills which will yield inaccurate result [9]. Moreover, it will be more challenging as the number of sample increases [10]. However, there is hematology counter in the market which is automated, fast and accurate but it is expensive [11]. In this work, the same objective is made but with a low cost approach.

The approach that this work offers is by using Convolutional Neural Network (CNN) which apply deep learning technique. It is less complex than the conventional method which the image has to go through many process and steps before being classified [12]. Preprocessing and feature extraction has to be done for conventional method [13]. It is important to extract the suitable features as it gives a huge impact to the classification accuracy. While in CNN, the structure is specially built to tackle the image variation and feature extraction problem. The image need to fed for the model to study and learn the pattern of each classes [14].

Deep CNN also known to solve computer vision issues successfully such as object recognition, semantic segmentation, object detection and video analysis [15]. It is widely used and applied on such as heartbeats classification [16], road crack detection [17], segmentation of blood vessels in retina image, skin cancer and lung lesion [18]. Other than that, some researchers use CNN for dynamic scene deblurring [19]. Google Colaboratory or Google colab is used along with CNN as it provides server less Jupyter notebook and it is free to use [20]. Some of works use Google colab for video- based emotion recognition [21] and breast cancer identification [22]. While this paper focuses on WBC classification using CNN. However, many works related to CNN WBC classification done but limited to one dataset which no comparison in terms of its performance has been made.

In conclusion, this paper emphasizes on the study of WBC classification performance for various datasets as WBC analysis is undeniably important to monitor a patient's health condition. It is done by using CNN of deep learning and it is applied through Google Colab medium as it is fast and requires less time for data training process. Three datasets were used with three CNN pre-trained models and the result of training and validation accuracy is compared. Result will show the best model for each dataset and which dataset suits a certain model

best. The datasets used are Kaggle, IDB-2 and LISC. While pretrained models involved are VGG-16, Mobilenet and Resnet.

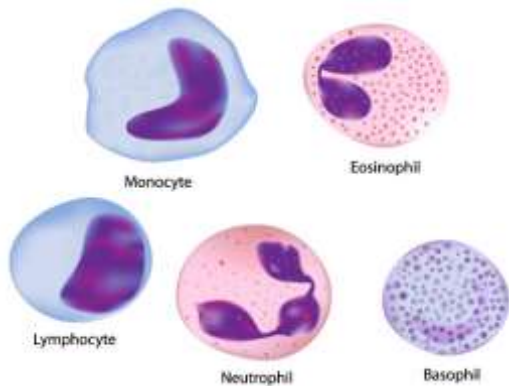


Fig. 1. Five types of WBC

This paper is organized as follows: Section II explains the flow of the method used. Section III shows the result and analysis. Lastly, Section IV concludes the result obtained and recommendations of future works.

II. SYSTEM OVERVIEW

A. Google Colab

In this paper, Google Colab is used as a platform to execute machine learning models in the cloud. It is a free Jupyter notebook and in Python. One of its advantages is it can be edited by the team members and easy to access without requiring any setup. It supported many types of machine learning libraries and can be easily loaded. As for this project, the flowchart of starting Google Colab is as depicted in Fig. 2.

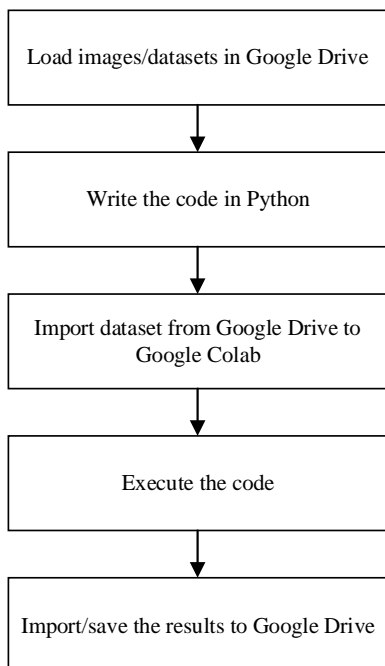


Fig. 2. Flowchart of Google Colab Settings.

Referring to the flowchart, setting up Google Colab is easy and does not require complicated steps to follow. Firstly, external dataset is uploaded to Google Drive. There are mainly two files for each dataset which are training and validation files. Before executing code in Google Colab, the dataset must be imported from Google Drive. Next, training and validation data directory is created. Pretrained model is constructed by integrating TensorFlow and Keras. After that, the training process is done and the outcome is plotted. The result is then saved in Google Drive file.

B. Convolutional Neural Network (CNN)

Varies of pretrained models from CNN is used in this project which are VGG-16, mobilenet and resnet. These models are different from one another but the main elements are the same which are convolution layer, non-linearity and fully connected layer [23]. Basically, images will be fed to convolution layer that act as a filter of different sizes for every model. The image size will be different after the convolution layer. The number of layer also differ depending on the model itself. Next, the features vector is minimized by applying non-linearity layer. Lastly, fully connected layer is assigned to classify the categories of the images.

- VGG-16: Structure of VGG-16 is considered simple as it consists only three main elements of convolution, max pooling and fully connected as shown in Fig. 3. Max pooling in VGG-16 is used to help with overfitting. Other than that, it reduces the number of parameters to learn which will reduce the computational cost. Overall, it has 13 convolution layers and 3 fully connected layers. Input image size for VGG-16 is fixed 224x224 RGB image.
- Mobilenet: Mobilenet is known as a small, low latency and low power model. It consists of 27 convolutional layers which includes depthwise convolution as depicted in Fig. 4. In mobilenet structure, initially, 3x3 convolutional layer is applied. Then followed by depthwise convolution layer and 1x1 convolution layer. These process is applied for 13 times as it has 13 depthwise convolution layers and 13 1x1 convolution layers. Next, average pool layer, fully connected and softmax layer is added to classify the classes of the image.

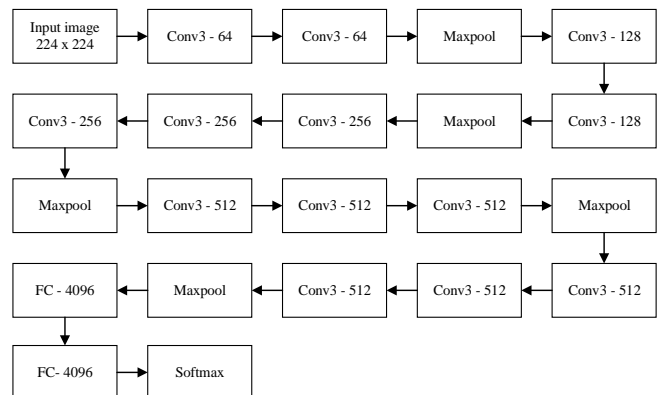


Fig. 3. VGG-16 Architecture.

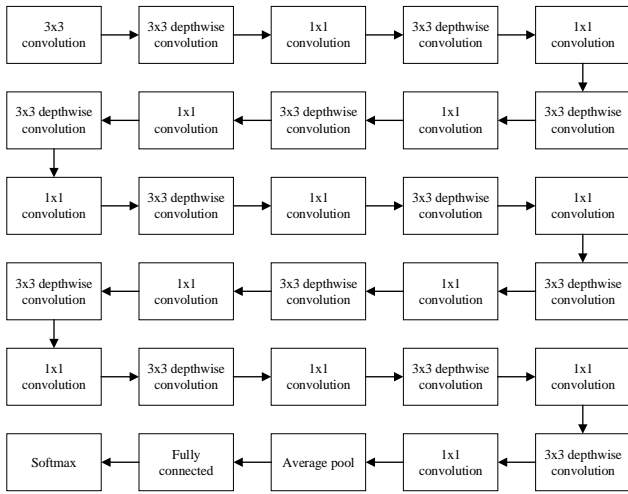


Fig. 4. Mobilenet Architecture.

- Resnet: Resnet architecture is acquired based on the Residual Network. It has 34-layer residual and this network implemented the skip connected as depicted in Fig. 5. Basically, it trains few layers and the output is connected directly. One of its advantages is any layer that can affect the performance of the network will be ignored and skipped by regularization. Hence, problems caused by vanishing or exploding gradient can be tackled by using this network.

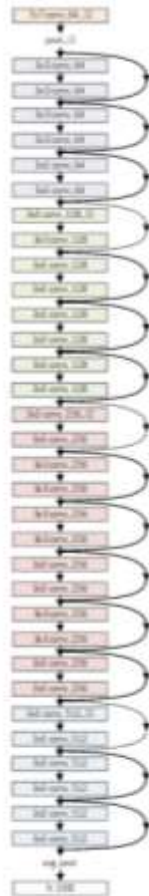


Fig. 5. Resnet Architecture.

C. Dataset

There are three datasets used in this paper which are Kaggle, IDB-2 and LISC.

1) *Kaggle*: In this database, there are four classes of white blood cell image which are Eosinophil, Neutrophil, Lymphocyte and Monocyte. The samples of images in this database is as shown in Fig. 6. There are 1500 images for each class and total image in the database is 6000 images for training. The images are in RGB and it has variation of image rotation.

2) *IDB-2*: There are only two classes which are Lymphoblast and Non-lymphoblast and the sample images are as depicted in Fig. 7. In this database, there are 130 images of lymphoblast and 130 of non-lymphoblast which makes the total image is 260. In this case, lymphoblast is an abnormal lymphocyte cells. The motive of using this dataset is to classify the lymphoblast and non-lymphoblast cell.

3) *LISC*: Images from this dataset is a healthy subject images that consists of five types of WBC which are Eosinophil, Neutrophil, Basophil, Lymphocyte and Monocyte. The sample image in LISC dataset is as shown in Fig. 8. The number of images in this dataset for Eosinophil, Neutrophil, Basophil, Lymphocyte and Monocyte is 39, 50, 53, 52 and 48 respectively. This dataset is different from the other two datasets by its image magnification which make the image contains all cell types and particles such as white blood cell, red blood cell and platelet. While the Kaggle and IDB focus more on the WBC region.

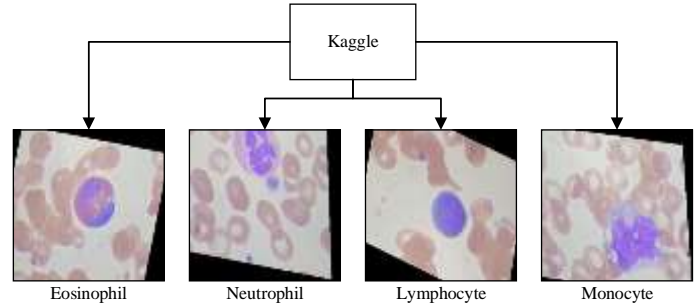


Fig. 6. Sample Images in Kaggle.

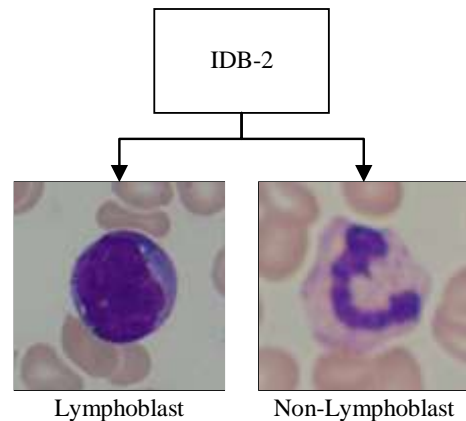


Fig. 7. Sample Images in IDB-2.

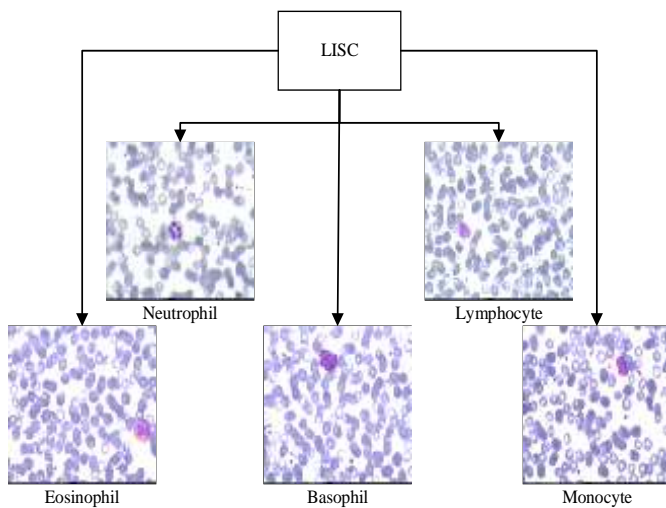


Fig. 8. Sample Images in LISC.

III. RESULT AND ANALYSIS

In this paper, three CNN models were tested using three different datasets to prove which model suits which dataset the best. Each dataset of Kaggle, IDB2 and LISC is trained using three models which are mobilenet, resnet and VGG-16. 70% of the images in each dataset were used for training and remaining 30% used for validation. Total number of image in IDB2 is 260 images and 182 images were used for training and 78 images were used for validation. While for LISC dataset, 168 images were used for training and 74 images were used for validation purpose. Epoch is fixed to 50 and the batch size is 64 in training process.

A. Dataset

Each dataset is tested using three different models of mobilenet, resnet and VGG-16. The results obtained is compared.

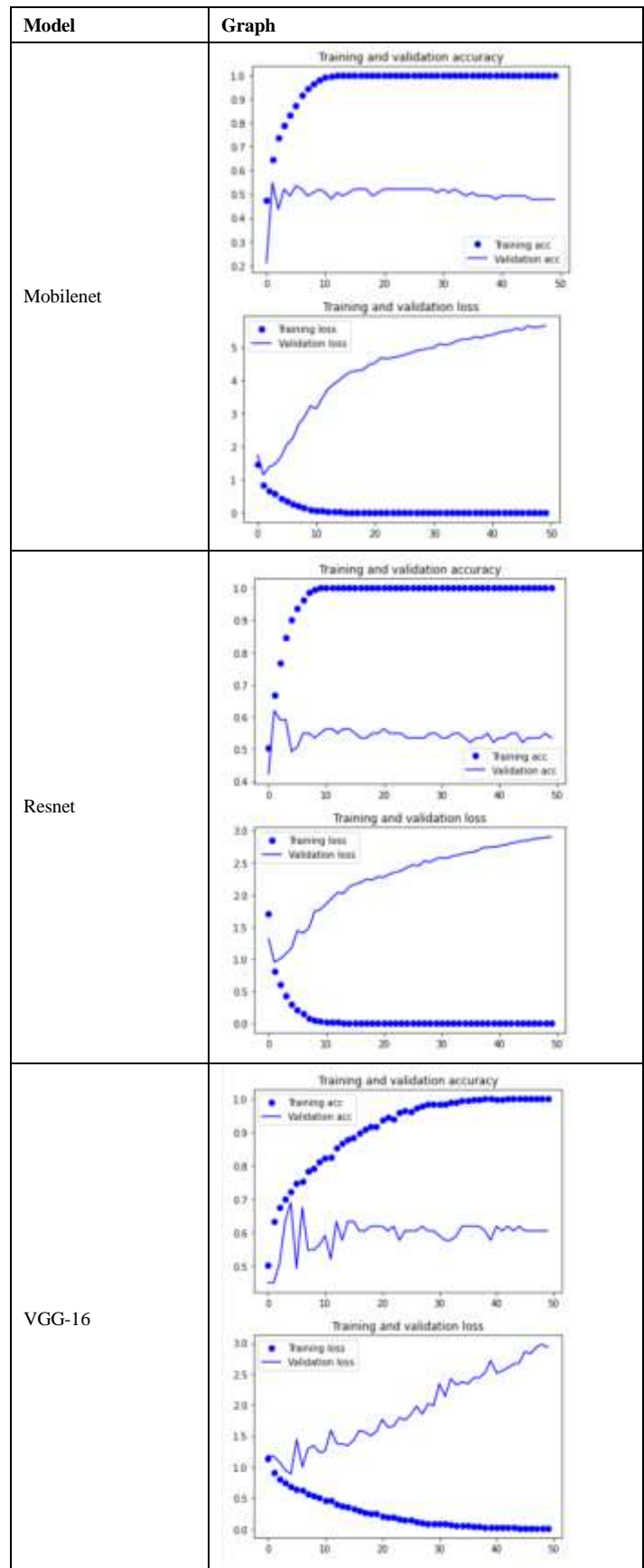
- Kaggle: As mentioned before, there are total of 6000 images in this dataset. It is trained using three models and the result is as tabulated in Table I.

Based on the table above, it can be seen that for training, the highest accuracy can be obtained by Resnet with lowest training loss. While for validation, highest accuracy is VGG-16 with lowest validation loss of 1.9219. Table II visualizes the graph plotting for both training and validation accuracy and loss. It can be seen that Resnet training accuracy pattern is more stable than the other two models. While for validation, the highest accuracy obtained is 0.6901 by VGG-16 model compared to the other two models, which barely obtained 0.6500 accuracy.

TABLE I. KAGGLE RESULT

Model \ Result	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
Mobilenet	0.1049	0.9629	4.3103	0.4983
Resnet	0.0884	0.9713	2.2724	0.5425
VGG-16	0.2547	0.9091	1.9219	0.5955

TABLE II. GRAPH OF TRAINING AND VALIDATION FOR KAGGLE



- IDB-2: In IDB-2, there are two classes of image which are Lymphoblast and Non-Lymphoblast cell. Both contains equal number of 130 images and the total for both classes is 260 images. All the images undergo all three models to get the result for comparison.

Based on the Table III, highest training accuracy of 0.9721 is obtained by VGG-16 model. It is the same for validation which achieved 0.7913 validation accuracy. The lowest accuracy for training is mobilenet by 0.0025 and for validation, the lowest is 0.6999 which obtained by resnet model. Next, Table IV compares the graph plotting for training and validation.

It can be seen that for training accuracy, VGG-16 is not the most stable but it increases gradually and its average accuracy is the highest. While for validation, it is clearly can be seen that VGG-16 can achieve more than 0.8 accuracy compared to mobilenet and resnet. VGG-16 is the most suitable model for IDB-2 database.

- LISC: LISC has the most classes consists of Basophil, Eosinophil, Neutrophil, Lymphocyte and Monocyte. The result for model comparison is as tabulated in Table V.

Resnet achieved the highest training accuracy of 0.9771 and lowest loss which is 0.0819. However, for validation, the highest accuracy is 0.5781 achieved by VGG-16. The validation accuracy for LISC database is not high and satisfactory. Table VI compares the graph for training and validation for each model of mobilenet, resnet and VGG-16. It can be seen that the consistency of training accuracy for resnet is better compared to mobilenet and VGG-16. However, it is different for validation accuracy where it fluctuated and not stable.

B. Model

This section explains the performance comparison of each dataset by model. Results showed the best dataset for which model.

- Mobilenet: In mobilenet, there are 27 convolutional layers which consist of 13 depthwise layers, 1 3x3 convolution layer and 13 1x1 convolution layers. This model is small, low latency and low power models.

Three datasets have been tested using this model for classification purposes and the result is as tabulated in Table VII.

Based on this result, it can clearly be seen that highest training accuracy is achieved on IDB-2 dataset by 96.96% and same applied to validation accuracy which 0.7210 is achieved for IDB-2. It can be said that IDB dataset is the most suitable dataset for mobilenet.

TABLE III. IDB-2 RESULT

Model \ Result	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
Mobilenet	0.1779	0.9696	1.0633	0.7210
Resnet	0.2286	0.9718	1.3799	0.6999
VGG-16	0.1279	0.9721	0.6396	0.7913

TABLE IV. GRAPH OF TRAINING AND VALIDATION FOR IDB-2

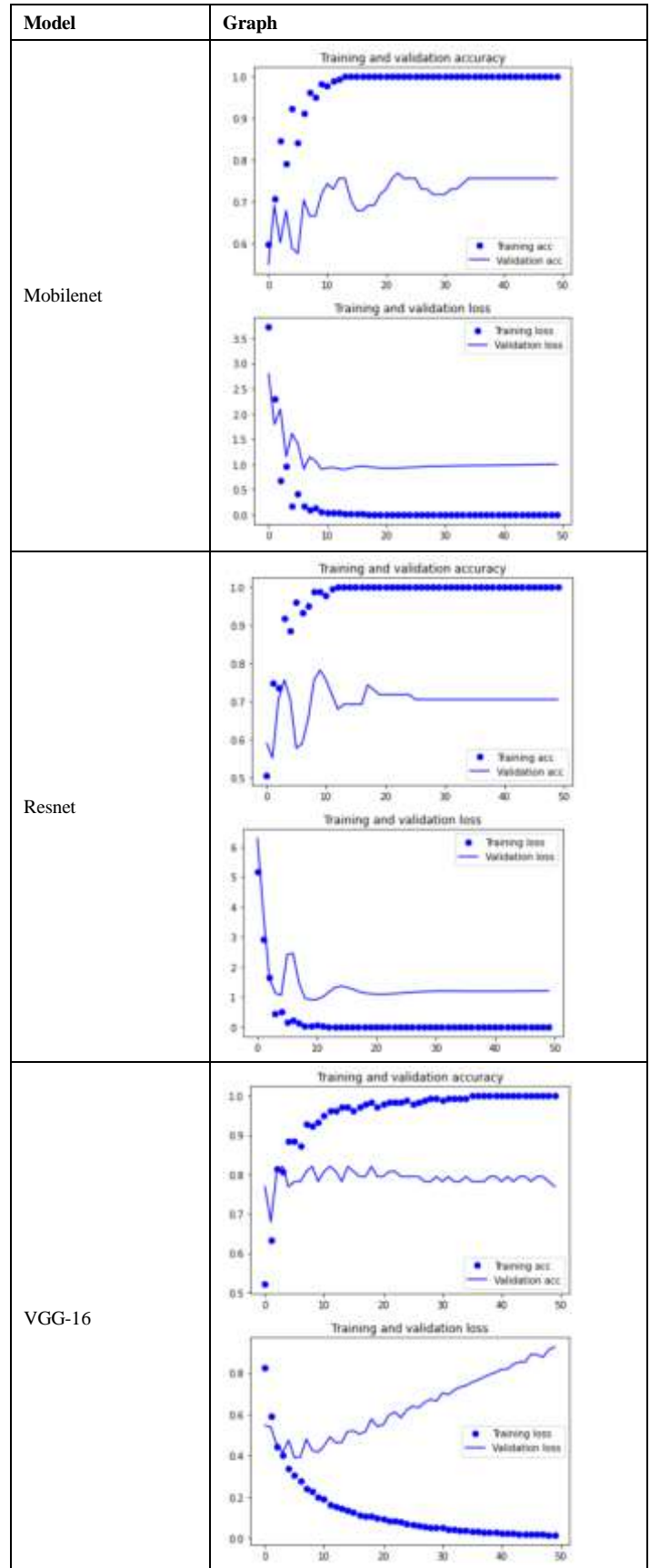


TABLE V. LISC RESULT

Result / Model	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
Mobilenet	0.3595	0.9413	2.0149	0.5778
Resnet	0.0819	0.9771	2.0235	0.5370
VGG-16	0.4332	0.8849	1.3299	0.5781

TABLE VI. GRAPH OF TRAINING AND VALIDATION FOR IDB-2

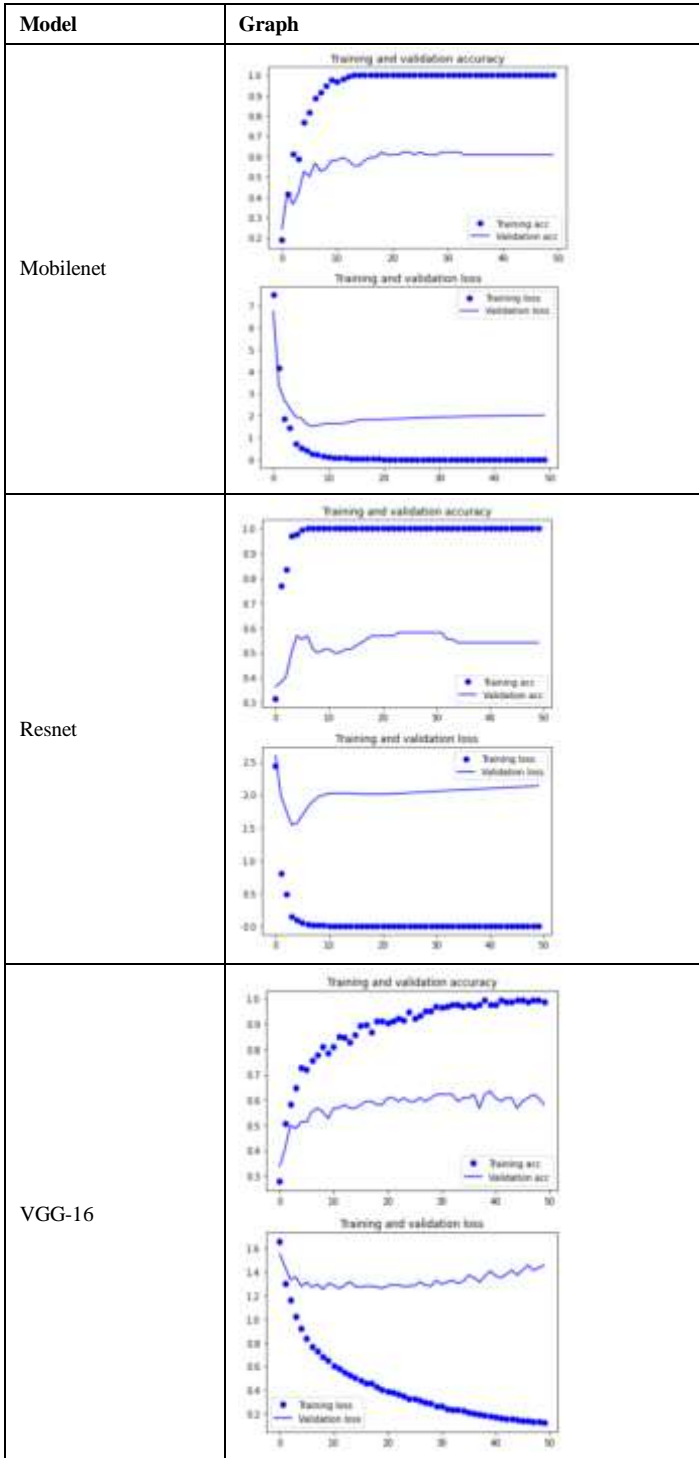


TABLE VII. DATASET COMPARISON FOR MOBILENET

Result / Model	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
Kaggle	0.1049	0.9629	4.3103	0.4983
IDB-2	0.1779	0.9696	1.0633	0.7210
LISC	0.3595	0.9413	2.0149	0.5778

- Resnet: Resnet is taken from Residual Network which means it contains residual layers. It has 34 residual layers and it also implement skip connection. The advantage of skip connection is to train without problems caused by vanishing and exploding gradient. The best dataset for resnet model is as shown in Table VIII.

TABLE VIII. DATASET COMPARISON FOR RESNET

Result / Model	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
Kaggle	0.0884	0.9713	2.2724	0.5425
IDB-2	0.2286	0.9718	1.3799	0.6999
LISC	0.0819	0.9771	2.0235	0.5370

As for training accuracy, the highest is obtained by LISC database. However, the differences between all the three datasets are not huge. While highest validation accuracy is achieved on IDB-2 dataset which is 0.6999.

- VGG-16: This model contains 16 layers which consist of 13 convolution layers and 3 fully connected layers. It is the simplest model compared to the other two models. The comparison for each dataset using VGG-16 is as tabulated in Table IX.

TABLE IX. DATASET COMPARISON FOR VGG-16

Result / Model	Training		Validation	
	Loss	Accuracy	Loss	Accuracy
Kaggle	0.2547	0.9091	1.9219	0.5955
IDB-2	0.1279	0.9721	0.6396	0.7913
LISC	0.4332	0.8849	1.3299	0.5781

It can be seen that IDB-2 dataset able to obtain highest training and validation accuracy of 0.9721 and 0.7913 respectively. Lowest accuracy for training and validation for VGG-16 is by LISC dataset.

IV. CONCLUSION AND FUTURE WORKS

This paper concludes the comparison of three models using three different databases to classify different classes. Google Colab was used for this project as it is fast, and can be edited by the team members. Other than that, it is free and supports many machine learning such as CNN. The CNN models involved in this paper are Mobilenet, Resnet and VGG-16. All the models tested using three different datasets of Kaggle, IDB-2 and LISC which contains 6000, 260 and 242 images respectively. Kaggle consists of four classes, IDB-2 consists of two classes and LISC consists of five classes of WBC types.

Firstly, Kaggle dataset is trained and validated using mobilenet, resnet and VGG-16 to know which model suits kaggle the best. It has four classes of data which are

Eosinophil, Neutrophil, Lymphocyte and Monocyte. Based on the experiment, resnet able to achieve highest training accuracy by 0.9713. However, VGG-16 is the highest for validation accuracy. But the differences between resnet and VGG-16 is 0.053. It can be said that resnet is the best model to classification of kaggle dataset. Next, IDB-2 database consists of two classes which are Lymphoblast and Non-Lymphoblast. The models are expected to detect and classify these two type of cells. Both training and validation accuracy is highest with VGG-16 model for IDB-2. Training accuracy achieved is 0.9721 and validation accuracy is 0.7913. It is clearly seen that VGG-16 is the best model to classify IDB-2 database. Last database tested is LISC database. It has five classes of data which are Basophil, Eosinophil, Neutrophil, Lymphocyte and Monocyte. In this case, highest training accuracy is obtained by resnet while highest validation accuracy is achieved by VGG-16 which is 0.9771 and 0.5781 respectively. The difference of validation accuracy between resnet and VGG-16 is 0.0411 which is not huge. As for LISC database, resnet is the best model to classify five classes of LISC database.

Next, the result is also manipulated by comparing the dataset result for each model. Firstly, for mobilenet, highest training and validation accuracy is from IDB-2 database which is 0.9696 and 0.7210 respectively. While for resnet, highest training accuracy is from LISC dataset. However, the differences between LISC and IDB-2 is only 0.0053. But for validation accuracy, IDB-2 is the highest which is 0.6999. Lastly, both training and validation accuracy is highest with IDB-2 database for VGG-16 model. The training accuracy is 0.9721 and the validation accuracy is 0.7913.

As a conclusion, resnet works best for kaggle and LISC dataset but VGG-16 works best for IDB-2 dataset. Other than that, the best dataset that can work with each model of mobilenet, resnet and VGG-16 is IDB-2 as it is less complex and it only has two classes of data. The images are also focused on the region of interest compared to the other two datasets which have more elements in the blood image.

In future, the own model is expected to be built which is better than the existing model to improve the training and validation accuracy. Next, other than classification, the system is also expected to localize the referred cell to make sure it is classifying the correct region. Experts' validation need to be obtained to strengthen the result's justification. Lastly, the number of dataset should be increased and undergo the same pretrained model.

ACKNOWLEDGMENT

This research is supported by a Fundamental Research Grant Scheme (FRGS) FRGS/1/2019/TK04/UTHM/02/6 (Vot. no K191) sponsored by the Ministry of Higher Education (MOHE). The research also received funding from the Office for Research, Innovation, Commercialization and Consultancy Management (ORICC), UTHM under Postgraduate Research Grant GPPS (Vot. No. H400).

REFERENCES

[1] Wang, X., et al., White blood cell counting on smartphone paper electrochemical sensor. *Biosensors and Bioelectronics*, 2017. 90: p. 549-557.

[2] Wang, Q., et al., A spectral and morphologic method for white blood cell classification. *Optics & Laser Technology*, 2016. 84: p. 144-148.

[3] Manik, S., L.M. Saini, and N. Vadera. Counting and classification of white blood cell using artificial neural network (ANN). in 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES). 2016. IEEE.

[4] Macawile, M.J., et al. White blood cell classification and counting using convolutional neural network. in 2018 3rd International Conference on Control and Robotics Engineering (ICCRE). 2018. IEEE.

[5] Othman, M.Z., T.S. Mohammed, and A.B. Ali, Neural network classification of white blood cell using microscopic images. *Int. J. Adv. Comput. Sci. Appl.*, 2017. 8(5): p. 99-104.

[6] Throngnumchai, K., et al. Classification of White blood cell using Deep Convolutional Neural Network. in 2019 12th Biomedical Engineering International Conference (BMEiCON). 2019. IEEE.

[7] Zheng, X., et al., Fast and robust segmentation of white blood cell images by self-supervised learning. *Micron*, 2018. 107: p. 55-71.

[8] Sona, K., C. Sriragavi, And A.V.B. Varshini, Detection Of White Blood Sample Cells Using Cnn. 2019.

[9] Habibzadeh, M., et al. Automatic white blood cell classification using pre-trained deep learning models: Resnet and inception. in Tenth International Conference on Machine Vision (ICMV 2017). 2018. International Society for Optics and Photonics.

[10] Safuan, S.N.M., M.R.M. Tomari, and W.N.W. Zakaria, White blood cell (WBC) counting analysis in blood smear images using various color segmentation methods. *Measurement*, 2018. 116: p. 543-555.

[11] Safuan, S.N.M., et al. Lymphoblast cell morphology identification to detect Acute Lymphoblastic Leukemia (ALL) using various color segmentation. in *Journal of Physics: Conference Series*. 2020. IOP Publishing.

[12] Dong, N., et al., White blood cell classification. *arXiv preprint arXiv:2008.07181*, 2020.

[13] Andrade, A.R., et al., Recent computational methods for white blood cell nuclei segmentation: A comparative study. *Computer Methods and Programs in Biomedicine*, 2019. 173: p. 1-14.

[14] Wang, R., et al. Convolutional recurrent neural networks for text classification. in 2019 International Joint Conference on Neural Networks (IJCNN). 2019. IEEE.

[15] Zhao, C., et al. Variational convolutional neural network pruning. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

[16] Acharya, U.R., et al., A deep convolutional neural network model to classify heartbeats. *Computers in biology and medicine*, 2017. 89: p. 389-396.

[17] Zhang, L., et al. Road crack detection using deep convolutional neural network. in 2016 IEEE international conference on image processing (ICIP). 2016. IEEE.

[18] Alom, M.Z., et al., Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv preprint arXiv:1802.06955*, 2018.

[19] Nah, S., T. Hyun Kim, and K. Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

[20] Bisong, E., Google Colaboratory, in *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. 2019, Springer. p. 59-64.

[21] Gunawan, T.S., et al., Development of video-based emotion recognition using deep learning with Google Colab. *TELKOMNIKA*, 2020. 18(5): p. 2463-2471.

[22] Balaraman, S., Comparison of Classification Models for Breast Cancer Identification using Google Colab. *Preprints*, 2020.

[23] Albawi, S., T.A. Mohammed, and S. Al-Zawi. Understanding of a convolutional neural network. in 2017 International Conference on Engineering and Technology (ICET). 2017. IEEE.